

นพดล พิมพ์จันทร์. 2549. ผลของข้อมูลที่มีค่าผิดปกติจากกลุ่มต่อความแกร่งของสัมประสิทธิ์
สหสัมพันธ์. วิทยานิพนธ์ปริญญาสาธรรณสุขศาสตรมหาบัณฑิต สาขาวิชาชีวสถิติ บัณฑิตวิทยาลัย
มหาวิทยาลัยขอนแก่น. [ISBN 974-116-843-8]
อาจารย์ที่ปรึกษาวิทยานิพนธ์ : ผู้ช่วยศาสตราจารย์ ดร. จิราพร เขียวอยู่

บทคัดย่อ

การวิจัยครั้งนี้เป็นการวิจัยเชิงจำลอง มีวัตถุประสงค์เพื่อศึกษาความแกร่งของสัมประสิทธิ์สหสัมพันธ์เมื่อพบข้อมูลที่มีค่าผิดปกติจากกลุ่ม ของตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ จำนวน 4 แบบ คือ สหสัมพันธ์แบบเพียร์สัน สหสัมพันธ์แบบสเปียร์แมน สหสัมพันธ์แบบเคนดอลล์ และ สหสัมพันธ์แบบถ่วงน้ำหนัก ศึกษาความแกร่งของสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ต่างๆที่แตกต่างกันดังนี้ ขนาดตัวอย่างมี 3 ระดับ คือ 20, 50 และ 100 กำหนดระดับความสัมพันธ์ที่ใช้ในการทดสอบ 5 ระดับคือ 0, 0.20, 0.50, 0.80 และ 1.0 กำหนดจำนวนของค่าผิดปกติจากกลุ่มซึ่งเป็นค่าผิดปกติจากกลุ่มระดับปานกลางด้านมีค่ามาก ให้มีจำนวนที่แตกต่างกันอยู่ในช่วง 0%, 5%, 10%, 20% และ 30% ของขนาดตัวอย่าง ตามลำดับ โดยแยกศึกษาการเกิดค่าผิดปกติจากกลุ่มที่ละตัวแปร ระดับนัยสำคัญในการทดสอบคือ 0.05 ทำการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล พัฒนาโปรแกรมจำลองแบบด้วยภาษาซีชาร์ป (C#) ในแต่ละสถานการณ์ทำการทดลองซ้ำ 1,000 ครั้งเกณฑ์ที่ใช้ในการเปรียบเทียบความแกร่งของสัมประสิทธิ์สหสัมพันธ์มี 2 แบบ คือ 1) ความแกร่งในการทดสอบทางสถิติ ได้แก่ ความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 อำนาจการทดสอบ และ ค่าประสิทธิภาพสัมพัทธ์ 2) ความแกร่งในการประมาณค่า ได้แก่ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ผลการศึกษาพบว่า เมื่อไม่มีค่าผิดปกติจากกลุ่ม และตัวอย่างมีขนาดเล็ก ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนมีความแกร่งที่สุด แต่ถ้าตัวอย่างมีขนาดใหญ่ตัวประมาณค่าสหสัมพันธ์แบบเพียร์สันมีความแกร่งที่สุด เมื่อมีค่าผิดปกติจากกลุ่มไม่เกิน 10% ของขนาดตัวอย่าง ตัวประมาณค่าสหสัมพันธ์แบบถ่วงน้ำหนักและตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน มีความแกร่งใกล้เคียงกัน ถ้าค่าผิดปกติจากกลุ่มมากกว่า 10% ของขนาดตัวอย่าง ตัวประมาณค่าสหสัมพันธ์แบบสเปียร์แมนจะมีความแกร่งที่สุด ดังนั้น ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์และตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนัก เป็นตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่มีความแกร่งเมื่อมีค่าผิดปกติจากกลุ่มในตัวอย่าง

Noppadol Pimchan. 2006. *Effect of Outliers on the Robustness of Correlation Coefficients*.

Master of Public Health Thesis in Biostatistics, Graduate School, Khon Kaen University.

[ISBN 974-116-843-8]

Thesis Advisor: Asst. Prof. Dr. Jiraporn Khiewyoo

ABSTRACT

This simulation research aimed to investigate the robustness of correlation coefficients with detected outliers. Four estimators of correlation coefficient were considered in the study. These included Pearson product moment correlation coefficient (r_{xy}), Spearman rank correlation coefficient (r_s), Kendall rank correlation coefficient (τ) and Biweight Midcorrelation (r_b). Robustness of statistical test and estimation were determined. The simulation was performed under all combination of following conditions. A first condition was 3 levels of sample size; n equal 20, 50, and 100 respectively. A second condition was 5 levels of strength of association; ρ equal 0.0, 0.20, 0.50, 0.80, and 1.00 respectively. A third condition was 5 levels, 0.0, 0.05, 0.10, 0.20 and 0.30 of proportion of positive outliers in sample on both variable X and Y. The data were generated through simulation using Monte Carlo technique and the experiment was repeated 1,000 times for each situation. The program for the simulation was developed by C# language. Indicators for assessing the robustness of statistical test were ability in controlling for type I error, power of test and relative efficiency. Indicator for assessing the robustness of estimator was mean square error. The finding shown that in case of no outlier with small sample size ($n=20$), the robust correlation coefficient was r_s , but with large sample size, the robust correlation coefficient was r_{xy} . In case of outliers occurred less than 10% of samples, the r_b and r_s had the same robustness properties but when outliers occurred more than 10% of sample r_s was the robust correlation coefficient. In conclusion, when there were outliers in sample both r_s and r_b were robust correlation coefficients.